arXiv:0912.1660v1 [math.OC] 9 Dec 2009

# GRADIENT-BASED METHODS FOR SPARSE RECOVERY *

WILLIAM W. HAGER†, DZUNG T. PHAN‡, AND HONGCHAO ZHANG§

**Abstract.** The convergence rate is analyzed for the SpaSRA algorithm (Sparse Reconstruction by Separable Approximation) for minimizing a sum $f(\mathbf{x}) + \psi(\mathbf{x})$ where $f$ is smooth and $\psi$ is convex, but possibly nonsmooth. It is shown that if $f$ is convex, then the error in the objective function at iteration $k$, for $k$ sufficiently large, is bounded by $a/(b+k)$ for suitable choices of $a$ and $b$. Moreover, if the objective function is strongly convex, then the convergence is $R$-linear. An improved version of the algorithm based on a cycle version of the BB iteration and an adaptive line search is given. The performance of the algorithm is investigated using applications in the areas of signal processing and image reconstruction.

**AMS subject classifications.** 90C06, 90C25, 65Y20, 94A08

**Key words.** SpaRSA, ISTA, sparse recovery, sublinear convergence, linear convergence, image reconstruction, denoising, compressed sensing, nonsmooth optimization, nonmonotone convergence, BB method

## 1. Introduction.

In this paper we consider the following optimization problem

$$\min_{\mathbf{x}\in\mathbb{R}^n} \quad \phi(\mathbf{x}) := f(\mathbf{x}) + \psi(\mathbf{x}), \tag{1.1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a smooth function, and $\psi : \mathbb{R}^n \to \mathbb{R}$ is convex. The function $\psi$, usually called the *regularizer* or *regularization function*, is finite for all $\mathbf{x} \in \mathbb{R}^n$, but possibly nonsmooth. An important application of (1.1), found in the signal processing literature, is the well-known $\ell_2 - \ell_1$ problem (called *basis pursuit denoising* in [7])

$$\min_{\mathbf{x}\in\mathbb{R}^n} \quad \frac{1}{2}\|\mathbf{A}\mathbf{x}-\mathbf{b}\|_2^2 + \tau\|\mathbf{x}\|_1, \tag{1.2}$$

where $\mathbf{A} \in \mathbb{R}^{k\times n}$ (usually $k \le n$), $\mathbf{b} \in \mathbb{R}^k$, $\tau \in \mathbb{R}$, $\tau \ge 0$, and $\|\cdot\|_1$ is the 1-norm.

Recently, Wright, Nowak, and Figueiredo [24] introduced the Sparse Reconstruction by Separable Approximation algorithm (SpaRSA) for solving (1.1). The algorithm has been shown to work well in practice. In [24] the authors establish global convergence of SpaRSA. In this paper, we prove an estimate of the form $a/(b+k)$ for the error in the objective function when $f$ is convex. If the objective function is strongly convex, then the convergence of the objective function and the iterates is at least R-linear. A strategy is presented for improving the performance of SpaRSA based on a cyclic Barzilai-Borwein step [8, 9, 13, 19] and an adaptive choice [15] for the reference function value in the line search. The paper concludes with a series of numerical experiments in the areas of signal processing and image reconstruction.

Throughout the paper $\nabla f(\mathbf{x})$ denotes the gradient of $f$, a row vector. The gradient of $f(\mathbf{x})$, arranged as a column vector, is $\mathbf{g}(\mathbf{x})$. The subscript $k$ often represents

* October 25, 2009. This material is based upon work supported by the National Science Foundation under Grant 0619080.

† hager@math.ufl.edu, http://www.math.ufl.edu/∼hager, PO Box 118105, Department of Mathematics, University of Florida, Gainesville, FL 32611-8105. Phone (352) 392-0281. Fax (352) 392-8357.

‡ dphan@math.ufl.edu, http://www.math.ufl.edu/∼dphan, PO Box 118105, Department of Mathematics, University of Florida, Gainesville, FL 32611-8105. Phone (352) 392-0281. Fax (352) 392-8357.

§ hozhang@math.lsu.edu, http://www.math.lsu.edu/∼hozhang, Department of Mathematics, 140 Lockett Hall, Center for Computation and Technology, Louisiana State University, Baton Rouge, LA 70803-4918. Phone (225) 578-1982. Fax (225) 578-4276.

the iteration number in an algorithm, and $\mathbf{g}_k$ stands for $\mathbf{g}(\mathbf{x}_k)$. $\|\cdot\|$ denotes $\|\cdot\|_2$, the Euclidean norm. $\partial\psi(\mathbf{y})$ is the subdifferential at $\mathbf{y}$, a set of row vectors. If $\mathbf{p} \in \partial\psi(\mathbf{y})$, then

$$\psi(\mathbf{x}) \geq \psi(\mathbf{y}) + \mathbf{p}(\mathbf{x} - \mathbf{y})$$

for all $\mathbf{x} \in \mathbb{R}^n$.

**2. The SpaRSA algorithm.** The SpaRSA algorithm, as presented in [24], is as follows:

Sparse Reconstruction by Separable Approximation (SpaRSA)

Given $\eta > 1$, $\sigma \in (0, 1)$, $[\alpha_{\min}, \alpha_{\max}] \subset (0, \infty)$, and starting guess $\mathbf{x}_1$.
Set $k = 1$.
Step 1. Choose $\alpha_0 \in [\alpha_{\min}, \alpha_{\max}]$
Step 2. Set $\alpha = \eta^j \alpha_0$ where $j \geq 0$ is the smallest integer such that

$$\phi(\mathbf{x}_{k+1}) \leq \phi_k^R - \sigma\alpha\|\mathbf{x}_{k+1} - \mathbf{x}_k\|^2 \text{ where}$$

$$\mathbf{x}_{k+1} = \arg\min\{\nabla f(\mathbf{x}_k)\mathbf{z} + \alpha\|\mathbf{z} - \mathbf{x}_k\|^2 + \psi(\mathbf{z}) : \mathbf{z} \in \mathbb{R}^n\}.$$

Step 3. If $\mathbf{x}_{k+1} = \mathbf{x}_k$, terminate.
Step 4. Set $k = k + 1$ and go to step 1.

The parameter $\alpha_0$ in [24] was taken to be the BB parameter [1] with safeguards:

$$\alpha_0 = \alpha_k^{BB} = \min \ \{\|\alpha\mathbf{s}_k - \mathbf{y}_k\| : \alpha_{\min} \leq \alpha \leq \alpha_{\max}\} \tag{2.1}$$

where $\mathbf{s}_k = \mathbf{x}_k - \mathbf{x}_{k-1}$ and $\mathbf{y}_k = \mathbf{g}_k - \mathbf{g}_{k-1}$. Also, in [24], the reference value $\phi_k^R$ is the GLL [14] reference value $\phi_k^{\max}$ defined by

$$\phi_k^{\max} = \max\{\phi(\mathbf{x}_{k-j}) : 0 \leq j < \min(k, M)\}. \tag{2.2}$$

In other words, at iteration $k$, $\phi_k^{\max}$ is the maximum of the $M$ most recent values for the objective function. Note that if $\mathbf{x}_{k+1} = \mathbf{x}_k$, then

$$\mathbf{0} \in \nabla f(\mathbf{x}_k) + \partial\psi(\mathbf{x}_{k+1}) = \nabla f(\mathbf{x}_{k+1}) + \partial\psi(\mathbf{x}_{k+1}).$$

Hence, $\mathbf{x}_{k+1} = \mathbf{x}_k$ is a stationary point.

The overall structure of the SpaRSA algorithm is closely related to that of the Iterative Shrinkage Thresholding Algorithm (ISTA) [6, 10, 12, 16, 23]. ISTA, however, employs a fixed choice for $\alpha$ related to the Lipschitz constant for $f$, while SpaRSA employs a nonmonotone line search. A sublinear convergence result for a monotone line search version of ISTA is given by Beck and Teboulle [2] and by Nesterov [18]. In Section 3 we give a sublinear convergence result for the nonmonotone SpaRSA, while Section 4 gives a linear convergence result when the objective function is strongly convex.

In [24] it is shown that the line search in Step 2 terminates for a finite $j$ when $f$ is Lipschitz continuously differentiable. Here we weaken this condition by only requiring Lipschitz continuity over a bounded set.

Proposition 2.1. *Let $\mathcal{L}$ be the level set defined by*

$$\mathcal{L} = \{\mathbf{x} \in \mathbb{R}^n : \phi(\mathbf{x}) \leq \phi(\mathbf{x}_1)\}. \tag{2.3}$$

*We make the following assumptions:*

(A1) *The level set $\mathcal{L}$ is contained in the interior of a compact, convex set $\mathcal{K}$, and $f$ is Lipschitz continuously differentiable on $\mathcal{K}$.*
(A2) *$\psi$ is convex and $\psi(\mathbf{x})$ is finite for all $\mathbf{x} \in \mathbb{R}^n$.*
*If $\phi(\mathbf{x}_k) \le \phi_k^R \le \phi(\mathbf{x}_1)$, then there exists $\bar{\alpha}$ with the property that*

$$\phi(\mathbf{x}_{k+1}) \le \phi_k^R - \sigma\alpha\|\mathbf{x}_{k+1} - \mathbf{x}_k\|^2$$

*whenever $\alpha \ge \bar{\alpha}$ where $\mathbf{x}_{k+1}$ is obtained as in Step 2 of SpaRSA.*

*Proof.* Let $\Phi_k$ be defined by

$$\Phi_k(\mathbf{z}) = f(\mathbf{x}_k) + \nabla f(\mathbf{x}_k)(\mathbf{z} - \mathbf{x}_k) + \alpha\|\mathbf{z} - \mathbf{x}_k\|^2 + \psi(\mathbf{z}),$$

where $\alpha \ge 0$. Since $\Phi_k$ is a strongly convex quadratic, its level sets are compact, and the minimizer $\mathbf{x}_{k+1}$ in Step 2 exists. Since $\mathbf{x}_{k+1}$ is the minimizer of $\Phi_k$, we have

$$\begin{aligned}\Phi_k(\mathbf{x}_{k+1}) &= f(\mathbf{x}_k) + \nabla f(\mathbf{x}_k)(\mathbf{x}_{k+1} - \mathbf{x}_k) + \alpha\|\mathbf{x}_{k+1} - \mathbf{x}_k\|^2 + \psi(\mathbf{x}_{k+1}) \\ &\le \Phi_k(\mathbf{x}_k) = f(\mathbf{x}_k) + \psi(\mathbf{x}_k).\end{aligned}$$

This is rearranged to obtain

$$\begin{aligned}\alpha\|\mathbf{x}_{k+1} - \mathbf{x}_k\|^2 &\le \nabla f(\mathbf{x}_k)(\mathbf{x}_k - \mathbf{x}_{k+1}) + \psi(\mathbf{x}_k) - \psi(\mathbf{x}_{k+1}) \\ &\le \nabla f(\mathbf{x}_k)(\mathbf{x}_k - \mathbf{x}_{k+1}) + \mathbf{p}_k(\mathbf{x}_k - \mathbf{x}_{k+1}),\end{aligned}$$

where $\mathbf{p}_k \in \partial\psi(\mathbf{x}_k)$. Taking norms yields

$$\|\mathbf{x}_{k+1} - \mathbf{x}_k\| \le (\|\mathbf{g}_k\| + \|\mathbf{p}_k\|)/\alpha. \tag{2.4}$$

By Theorem 23.4 and Corollary 24.5.1 in [20] and by the compactness of $\mathcal{L}$, there exists a constant $c$, independent of $\mathbf{x}_k \in \mathcal{L}$, such that $\|\mathbf{g}_k\| + \|\mathbf{p}_k\| \le c$. Consequently, we have

$$\|\mathbf{x}_{k+1} - \mathbf{x}_k\| \le c/\alpha.$$

Since $\mathcal{K}$ is compact and $\mathcal{L}$ lies in the interior of $\mathcal{K}$, the distance $\delta$ from $\mathcal{L}$ to the boundary of $\mathcal{K}$ is positive. Choose $\beta \in (0, \infty)$ so that $c/\beta \le \delta$. Hence, when $\alpha \ge \beta$, $\mathbf{x}_{k+1} \in \mathcal{K}$ since $\mathbf{x}_k \in \mathcal{L}$.

Let $\lambda$ denote the Lipschitz constant for $f$ on $\mathcal{K}$ and suppose that $\alpha \ge \beta$. Since $\mathbf{x}_k \in \mathcal{L} \subset \mathcal{K}$ and $\|\mathbf{x}_{k+1} - \mathbf{x}_k\| \le \delta$, we have $\mathbf{x}_{k+1} \in \mathcal{K}$. Moreover, due to the convexity of $\mathcal{K}$, the line segment connecting $\mathbf{x}_k$ and $\mathbf{x}_{k+1}$ lies in $\mathcal{K}$. Proceeding as in [24], a Taylor expansion around $\mathbf{x}_k$ yields

$$f(\mathbf{x}_{k+1}) \le f(\mathbf{x}_k) + \nabla f(\mathbf{x}_k)(\mathbf{x}_{k+1} - \mathbf{x}_k) + .5\lambda\|\mathbf{x}_{k+1} - \mathbf{x}_k\|^2.$$

Adding $\psi(\mathbf{x}_{k+1})$ to both sides, we have

$$\begin{aligned}\phi(\mathbf{x}_{k+1}) &\le \Phi_k(\mathbf{x}_{k+1}) + (.5\lambda - \alpha)\|\mathbf{x}_{k+1} - \mathbf{x}_k\|^2 \\ &\le \Phi_k(\mathbf{x}_k) + (.5\lambda - \alpha)\,\|\mathbf{x}_{k+1} - \mathbf{x}_k\|^2 \\ &= \phi(\mathbf{x}_k) + (.5\lambda - \alpha)\,\|\mathbf{x}_{k+1} - \mathbf{x}_k\|^2 \\ &\le \phi_k^R + (.5\lambda - \alpha)\,\|\mathbf{x}_{k+1} - \mathbf{x}_k\|^2 \quad \text{since } \phi(\mathbf{x}_k) \le \phi_k^R \\ &\le \phi_k^R - \sigma\alpha\|\mathbf{x}_{k+1} - \mathbf{x}_k\|^2 \quad \text{if } .5\lambda - \alpha \le -\sigma\alpha.\end{aligned} \tag{2.5}$$

Hence, the proposition holds with

$$\bar{\alpha} = \max\left\{\beta, \frac{\lambda}{2(1-\sigma)}\right\}.$$

□

REMARK 1. *Suppose $\phi_k^R \le \phi(\mathbf{x}_1)$. In Step 2 of SpaRSA, $\mathbf{x}_{k+1}$ is chosen so that $\phi(\mathbf{x}_{k+1}) \le \phi_k^R$. Hence, there exists $\phi_{k+1}^R$ such that $\phi(\mathbf{x}_{k+1}) \le \phi_{k+1}^R \le \phi(\mathbf{x}_1)$. In other words, if the hypothesis "$\phi(\mathbf{x}_k) \le \phi_k^R \le \phi(\mathbf{x}_1)$" of Proposition 2.1 is satisfied at step $k$, then a choice for $\phi_{k+1}^R$ exists which satisfies this hypothesis at step $k+1$.*

REMARK 2. *We now show that the GLL reference value $\phi_k^{\max}$ satisfies the condition $\phi(\mathbf{x}_k) \le \phi_k^R \le \phi(\mathbf{x}_1)$ of Proposition 2.1 for each $k$. The condition $\phi_k^{\max} \ge \phi(\mathbf{x}_k)$ is a trivial consequence of the definition of $\phi_k^{\max}$. Also, by the definition, we have $\phi_1^{\max} = \phi(\mathbf{x}_1)$. For $k \ge 1$, $\phi(\mathbf{x}_{k+1}) \le \phi_k^{\max}$ according to Step 2 of SpaRSA. Hence, $\phi_k^{\max}$ is a decreasing function of $k$. In particular, $\phi_k^{\max} \le \phi_1^{\max} = \phi(\mathbf{x}_1)$.*

**3. Convergence estimate for convex functions.** In this section we give a sublinear convergence estimate for the error in the objective function value $\phi(\mathbf{x}_k)$ assuming $f$ is convex and the assumptions of Proposition 2.1 hold.

By (A1) and (A2), (1.1) has a solution $\mathbf{x}^* \in \mathcal{L}$ and an associated objective function value $\phi^* := \phi(\mathbf{x}^*)$. The convergence of the objective function values to $\phi^*$ is a consequence of the analysis in [24]:

LEMMA 3.1. *If* (A1) *and* (A2) *hold and $\phi_k^R = \phi_k^{\max}$ for every $k$, then*

$$\lim_{k\to\infty} \phi(\mathbf{x}_k) = \phi^*.$$

*Proof.* By [24, Lemma 4], the objective function values $\phi(\mathbf{x}_k)$ approach a limit denoted $\bar{\phi}$. By [24, Theorem 1], all accumulation points of the iterates $\mathbf{x}_k$ are stationary points. An accumulation point exists since $\mathcal{K}$ is compact and the iterates are all contained in $\mathcal{L} \subset \mathcal{K}$, as shown in Remark 2. Since $f$ and $\psi$ are both convex, a stationary point is a global minimizer of $\phi$. Hence, $\bar{\phi} = \phi^*$. □

Our sublinear convergence result is the following:

THEOREM 3.2. *If* (A1) *and* (A2) *hold, $f$ is convex, and $\phi_k^R = \phi_k^{\max}$ for all $k$, then there exist constants $a$ and $b$ such that*

$$\phi(\mathbf{x}_k) - \phi^* \le \frac{a}{b+k}$$

*for $k$ sufficiently large.*

*Proof.* By (2.5) with $k+1$ replaced by $k$, we have

$$\phi(\mathbf{x}_k) \le \Phi_{k-1}(\mathbf{x}_k) + b_0\|\mathbf{s}_k\|^2, \quad b_0 = .5\lambda, \tag{3.1}$$

where $\mathbf{s}_k = \mathbf{x}_k - \mathbf{x}_{k-1}$. Since $\mathbf{x}_k$ minimizes $\Phi_{k-1}$ and $f$ is convex, it follows that

$$\begin{aligned}
\Phi_{k-1}(\mathbf{x}_k) &= \min_{\mathbf{z}\in\mathbb{R}^n} \{f(\mathbf{x}_{k-1}) + \nabla f(\mathbf{x}_{k-1})(\mathbf{z}-\mathbf{x}_{k-1}) + \alpha_{k-1}\|\mathbf{z}-\mathbf{x}_{k-1}\|^2 + \psi(\mathbf{z})\} \\
&\le \min\{f(\mathbf{z}) + \psi(\mathbf{z}) + \alpha_{k-1}\|\mathbf{z}-\mathbf{x}_{k-1}\|^2 : \mathbf{z}\in\mathbb{R}^n\} \\
&= \min\{\phi(\mathbf{z}) + \alpha_{k-1}\|\mathbf{z}-\mathbf{x}_{k-1}\|^2 : \mathbf{z}\in\mathbb{R}^n\},
\end{aligned} \tag{3.2}$$

where $\alpha_{k-1}$ is the terminating value of $\alpha$ at step $k-1$. Combining (3.1) and (3.2) gives

$$\phi(\mathbf{x}_k) \le \min\{\phi(\mathbf{z}) + \bar{\beta}\|\mathbf{z} - \mathbf{x}_{k-1}\|^2 : \mathbf{z} \in \mathbb{R}^n\} + b_0\|\mathbf{s}_k\|^2, \tag{3.3}$$

where $\bar{\beta} = \eta\bar{\alpha}$ is an upper bound for the $\alpha_k$ implied by Proposition 2.1. By the convexity of $\phi$ and with $\mathbf{z} = (1-\lambda)\mathbf{x}_{k-1} + \lambda\mathbf{x}^*$ for any $\lambda \in [0, 1]$, we have

$$\begin{aligned}\min_{\mathbf{z}\in\mathbb{R}^n} \phi(\mathbf{z}) + \bar{\beta}\|\mathbf{z} - \mathbf{x}_{k-1}\|^2 &\le \phi((1-\lambda)\mathbf{x}_{k-1} + \lambda\mathbf{x}^*) + \bar{\beta}\lambda^2\|\mathbf{x}_{k-1} - \mathbf{x}^*\|^2 \\ &\le (1-\lambda)\phi(\mathbf{x}_{k-1}) + \lambda\phi^* + \bar{\beta}\lambda^2\|\mathbf{x}_{k-1} - \mathbf{x}^*\|^2 \\ &= (1-\lambda)\phi(\mathbf{x}_{k-1}) + \lambda\phi^* + b_k\lambda^2,\end{aligned}$$

where $b_k = \bar{\beta}\|\mathbf{x}_{k-1} - \mathbf{x}^*\|^2$. Combining this with (3.3) yields

$$\begin{aligned}\phi(\mathbf{x}_k) &\le (1-\lambda)\phi(\mathbf{x}_{k-1}) + \lambda\phi^* + b_k\lambda^2 + b_0\|\mathbf{s}_k\|^2 \\ &\le (1-\lambda)\phi^R_{k-1} + \lambda\phi^* + b_k\lambda^2 + b_0\|\mathbf{s}_k\|^2 \end{aligned} \tag{3.4}$$

for any $\lambda \in [0, 1]$. Define

$$\phi_i = \max\{\phi(\mathbf{x}_k) : (i-1)M < k \le iM\} = \phi^R_{iM}, \tag{3.5}$$

and let $k_i$ denote the index $k$ where the maximum is attained. Since $\phi(\mathbf{x}_{k+1}) \le \phi^R_k$ in Step 2 of SpaRSA, it follows that $\phi^R_k = \phi^{\max}_k$ is a nonincreasing function of $k$. By (3.4) with $k = k_i$ and by the monotonicity of $\phi^R_k$, we have

$$\phi_i \le (1-\lambda)\phi_{i-1} + \lambda\phi^* + b_{k_i}\lambda^2 + b_0\|\mathbf{s}_{k_i}\|^2 \tag{3.6}$$

for any $\lambda \in [0, 1]$. Since both $\mathbf{x}_{k-1}$ and $\mathbf{x}^*$ lie in $\mathcal{L}$, it follows that

$$b_k = \bar{\beta}\|\mathbf{x}_{k-1} - \mathbf{x}^*\|^2 \le \bar{\beta}(\text{diameter of } \mathcal{L})^2 := b_2 < \infty. \tag{3.7}$$

Step 2 of SpaRSA implies that

$$\|\mathbf{s}_k\|^2 \le (\phi^R_{k-1} - \phi(\mathbf{x}_k))/b_1$$

where $b_1 = \sigma\alpha_{\min}$. We take $k = k_i$ and again exploit the monotonicity of $\phi^R_k$ to obtain

$$\|\mathbf{s}_{k_i}\|^2 \le (\phi_{i-1} - \phi_i)/b_1. \tag{3.8}$$

Combining (3.6)–(3.8) gives

$$\phi_i \le (1-\lambda)\phi_{i-1} + \lambda\phi^* + b_2\lambda^2 + b_3(\phi_{i-1} - \phi_i), \quad b_3 = b_0/b_1, \tag{3.9}$$

for every $\lambda \in [0, 1]$, The minimum on the right side is attained with the choice

$$\lambda = \min\left\{1, \frac{\phi_{i-1} - \phi^*}{2b_2}\right\}. \tag{3.10}$$

As a consequence of Lemma 3.1, $\phi_{i-1}$ converges to $\phi^*$. Hence, the minimizing $\lambda$ also approaches 0 as $i$ tends to $\infty$. Choose $k$ large enough that the minimizing $\lambda$ is less than 1. It follows from (3.9) that for this minimizing choice of $\lambda$, we have

$$\phi_i \le \phi_{i-1} - \frac{(\phi_{i-1} - \phi^*)^2}{4b_2} + b_3(\phi_{i-1} - \phi_i). \tag{3.11}$$

Define $e_i = \phi_i - \phi^*$. Subtracting $\phi^*$ from each side of (3.11) gives

$$\begin{aligned} e_i &\le e_{i-1} - e_{i-1}^2/(4b_2) + b_3(e_{i-1} - e_i) \\ &= (1 + b_3)e_{i-1} - e_{i-1}^2/(4b_2) - b_3 e_i. \end{aligned}$$

We arrange this to obtain

$$e_i \le e_{i-1} - b_4 e_{i-1}^2 \quad \text{where } b_4 = \frac{1}{4b_2(1 + b_3)}. \tag{3.12}$$

By (3.12) $e_i \le e_{i-1}$, which implies that

$$e_i \le e_{i-1} - b_4 e_{i-1} e_i \quad \text{or} \quad e_i \le \frac{e_{i-1}}{1 + b_4 e_{i-1}}.$$

We form the reciprocal of this last inequality to obtain

$$\frac{1}{e_i} \ge \frac{1}{e_{i-1}} + b_4.$$

Applying this inequality recursively gives

$$\frac{1}{e_i} \ge \frac{1}{e_j} + (i - j)b_4 \quad \text{or} \quad e_i \le \frac{e_j}{1 + (i - j)b_4 e_j},$$

where $j$ is chosen large enough to ensure that the minimizing $\lambda$ in (3.10) is less than 1 for all $i \ge j$.

Suppose that $k \in ((i-1)M, iM]$ with $i > j$. Since $i \ge k/M$, we have

$$\phi(\mathbf{x}_k) - \phi^* \le e_i \le \frac{e_j}{1 + (i - j)b_4 e_j} \le \frac{e_j}{1 - jb_4 e_j + kb_4 e_j/M}.$$

The proof is completed by taking $a = M/b_4$ and $b = M/(b_4 e_j) - Mj$. ☐

**4. Convergence estimate for strongly convex functions.** In this section we prove that SpaRSA converges R-linearly when $f$ is a convex function and $\phi$ satisfies

$$\phi(\mathbf{y}) \ge \phi(\mathbf{x}^*) + \mu\|\mathbf{y} - \mathbf{x}^*\|^2 \tag{4.1}$$

for all $\mathbf{y} \in \mathbb{R}^n$, where $\mu > 0$. Hence, $\mathbf{x}^*$ is a unique minimizer of $\phi$. For example, if $f$ is a strongly convex function, then (4.1) holds.

THEOREM 4.1. *If* (A1) *and* (A2) *hold, $f$ is convex, $\phi$ satisfies (4.1), and $\phi_k^R = \phi_k^{\max}$ for every $k$, then there exist constants $\theta \in (0,1)$ and $c$ such that*

$$\phi(\mathbf{x}_k) - \phi^* \le c\theta^k(\phi(\mathbf{x}_1) - \phi^*) \tag{4.2}$$

*for every $k$.*

*Proof.* Let $\phi_i$ be defined as in (3.5). We will show that there exist $\gamma \in (0,1)$ such that

$$\phi_i - \phi^* \le \gamma(\phi_{i-1} - \phi^*). \tag{4.3}$$

Let $c_1$ be chosen to satisfy the inequality

$$0 < c_1 < \min\left\{\frac{1}{2b_0}, \frac{\mu}{4b_0\bar{\beta}}\right\}. \tag{4.4}$$

We consider 2 cases.

*Case* 1. $\|\mathbf{s}_{k_i}\|^2 \ge c_1(\phi_{i-1} - \phi^*)$.

By (3.8), we have

$$c_1(\phi_{i-1} - \phi^*) \le (\phi_{i-1} - \phi_i)/b_1.$$

This can be rearranged to obtain

$$\phi_i - \phi^* \le (1 - b_1 c_1)(\phi_{i-1} - \phi^*),$$

which yields (4.3).

*Case* 2. $\|\mathbf{s}_{k_i}\|^2 < c_1(\phi_{i-1} - \phi^*)$.

We utilize the inequality (3.6) but with different bounds for the $b_{k_i}$ and $\mathbf{s}_{k_i}$ terms. For $k \in ((i-1)M, iM]$, we have

$$\begin{aligned} b_k := \bar{\beta}\|\mathbf{x}_{k-1} - \mathbf{x}^*\|^2 \le \frac{\bar{\beta}}{\mu}(\phi(\mathbf{x}_{k-1}) - \phi^*) &\le \frac{\bar{\beta}}{\mu}(\phi^R_{k-1} - \phi^*) \\ &\le \frac{\bar{\beta}}{\mu}(\phi^R_{(i-1)M} - \phi^*) = b_5(\phi_{i-1} - \phi^*), \quad b_5 = \frac{\bar{\beta}}{\mu}. \end{aligned}$$

The first inequality is due to (4.1) and the last inequality is since $\phi^R_k$ is monotone decreasing. By the definition of $k_i$ below (3.5), it follows that $k_i \in ((i-1)M, iM]$ and

$$b_{k_i} \le b_5(\phi_{i-1} - \phi^*). \tag{4.5}$$

Inserting in (3.6) the bound (4.5) and the Case 2 requirement $\|\mathbf{s}_{k_i}\|^2 < c_1(\phi_{i-1} - \phi^*)$ yields

$$\phi_i \le (1-\lambda)\phi_{i-1} + \lambda\phi^* + b_5(\phi_{i-1} - \phi^*)\lambda^2 + b_0 c_1(\phi_{i-1} - \phi^*)$$

for all $\lambda \in [0, 1]$. Subtract $\phi^*$ from each side to obtain

$$e_i \le [1 + b_0 c_1 - \lambda + b_5\lambda^2]e_{i-1} \tag{4.6}$$

for all $\lambda \in [0, 1]$.

The $\lambda \in [0, 1]$ which minimizes the coefficient of $e_{i-1}$ in (4.6) is

$$\lambda = \min\left\{1, \frac{1}{2b_5}\right\}.$$

If the minimizing $\lambda$ is 1, then $b_5 \le 1/2$ and the minimizing coefficient in (4.6) is

$$\gamma = b_0 c_1 + b_5 \le b_0 c_1 + 1/2 < 1$$

since $c_1 < 1/(2b_0)$ by (4.4). On the other hand, if the minimizing $\lambda$ is less than 1, then $b_5 > 1/2$ and the minimizing coefficient is

$$\gamma = 1 + b_0 c_1 - \frac{1}{4b_5} < 1$$

since $1/(4b_5) = \mu/(4\bar{\beta}) > b_0 c_1$ by (4.4). This completes the proof of (4.3).

For $k \in ((i-1)M, iM]$, we have

$$\phi(\mathbf{x}_k) - \phi^* \le e_i \le \gamma^{i-1} e_1 \le \frac{1}{\gamma}\left(\gamma^{1/M}\right)^k (\phi(\mathbf{x}_1) - \phi^*).$$

Hence, (4.2) holds with $c = 1/\gamma$ and $\theta = \gamma^{1/M}$. This completes the proof. ∎

Remark 3. *The condition (4.1) when combined with (4.2) shows that the iterates $\mathbf{x}_k$ converge R-linearly to $\mathbf{x}^*$.*

**5. More general reference function values.** The GLL reference function value $\phi_k^{\max}$, defined in (2.2), often leads to greater efficiency when $M > 1$, when compared to the monotone choice $M = 1$. In practice, it is found that even more flexibility in the reference function value can further accelerate convergence. In [15] we prove convergence of the nonmonotone gradient projection method whenever the reference function $\phi_k^R$ satisfies the following conditions:

(R1) $\phi_1^R = \phi(\mathbf{x}_1)$.
(R2) $\phi(\mathbf{x}_k) \le \phi_k^R \le \max\{\phi_{k-1}^R, \phi_k^{\max}\}$ for each $k > 1$.
(R3) $\phi_k^R \le \phi_k^{\max}$ infinitely often.

In [15] we provide a specific choice for $\phi_k^R$ which satisfies (R1)–(R3) and which gave more rapid convergence than the choice $\phi_k^R = \phi_k^{\max}$. To satisfy (R3), we could choose an integer $L > 0$ and simply set $\phi_k^R = \phi_k^{\max}$ every $L$ iterations. Another strategy, closer in spirit to what is used in the numerical experiments, is to choose a decrease parameter $\Delta > 0$ and set $\phi_k^R = \phi_k^{\max}$ if $\phi(\mathbf{x}_{k-L}) - \phi(\mathbf{x}_k) \le \Delta$. We now give convergence results for SpaRSA whenever the reference function value satisfies (R1)–(R3). In the first convergence result which follows, convexity of $f$ is not required.

Theorem 5.1. *If* (A1) *and* (A2) *hold and the reference function value $\phi_k^R$ satisfies* (R1)–(R3)*, then the iterates $\mathbf{x}_k$ of SpaRSA have a subsequence converging to a limit $\bar{\mathbf{x}}$ satisfying $\mathbf{0} \in \partial\phi(\bar{\mathbf{x}})$.*

*Proof.* We first apply Proposition 2.1 to show that Step 2 of SpaRSA is fulfilled for some choice of $j$. This requires that we show $\phi_k^R \le \phi(\mathbf{x}_1)$ for each $k$. This holds for $k = 1$ by (R1). Also, for $k = 1$, we have $\phi_1^{\max} = \phi(\mathbf{x}_1)$. Proceeding by induction, suppose that $\phi_i^R \le \phi(\mathbf{x}_1)$ and $\phi_i^{\max} \le \phi(\mathbf{x}_1)$ for $i = 1, 2, \ldots, k$. By Proposition 2.1, Step 2 of SpaRSA terminates at a finite $j$ and hence,

$$\phi(\mathbf{x}_{k+1}) \le \phi_k^R \le \phi(\mathbf{x}_1).$$

It follows that $\phi_{k+1}^{\max} \le \phi(\mathbf{x}_1)$ and $\phi_{k+1}^R \le \max\{\phi_k^R, \phi_{k+1}^{\max}\} \le \phi(\mathbf{x}_1)$. This completes the induction step, and hence, by Proposition 2.1, it follows that in every iteration, Step 2 of SpaRSA is fulfilled for a finite $j$.

By Step 2 of SpaRSA, we have

$$\phi(\mathbf{x}_k) \le \phi_{k-1}^R - \sigma\alpha_{\min}\|\mathbf{s}_k\|^2,$$

where $\mathbf{s}_k = \mathbf{x}_k - \mathbf{x}_{k-1}$. In the third paragraph of the proof of Theorem 2.2 in [15], it is shown that when an inequality of this form is satisfied for a reference function value satisfying (R1)–(R3), then

$$\liminf_{k\to\infty} \|\mathbf{s}_k\| = 0.$$

Let $k_i$ denote a strictly increasing sequence with the property that $\mathbf{s}_{k_i}$ tends to $\mathbf{0}$ and $\mathbf{x}_{k_i}$ approaches a limit denoted $\bar{\mathbf{x}}$. That is,

$$\lim_{i\to\infty} \mathbf{s}_{k_i} = 0 \quad \text{and} \quad \lim_{i\to\infty} \mathbf{x}_{k_i} = \bar{\mathbf{x}}.$$

Since $\mathbf{s}_{k_i}$ tends to $\mathbf{0}$, it follows that $\mathbf{x}_{k_i-1}$ also approaches $\bar{\mathbf{x}}$. By the first-order optimality conditions for $\mathbf{x}_{k_i}$, we have

$$\mathbf{0} \in \nabla f(\mathbf{x}_{k_i-1}) + 2\alpha_{k_i}(\mathbf{x}_{k_i} - \mathbf{x}_{k_i-1}) + \partial\psi(\mathbf{x}_{k_i}), \tag{5.1}$$

where $\alpha_{k_i}$ denotes the value of $\alpha$ in Step 2 of SpaRSA associated with $\mathbf{x}_{k_i}$. Again, by Proposition 2.1, we have the uniform bound $\alpha_{k_i} \le \bar{\beta} = \eta\bar{\alpha}$. Taking the limit as $i$ tends to $\infty$, it follows from Corollary 24.5.1 in [20] that

$$\mathbf{0} \in \nabla f(\bar{\mathbf{x}}) + \partial\psi(\bar{\mathbf{x}}).$$

This completes the proof. □

With a small change in (R3), we obtain either sublinear or linear convergence of the entire iteration sequence.

Theorem 5.2. *Suppose that* (A1) *and* (A2) *hold, $f$ is convex, the reference function value $\phi_k^R$ satisfies* (R1) *and* (R2)*, and there is $L > 0$ with the property that for each $k$,*

$$\phi_j^R \le \phi_j^{\max} \quad \textit{for some } j \in [k, k+L). \tag{5.2}$$

*Then there exist constants $a$ and $b$ such that*

$$\phi(\mathbf{x}_k) - \phi^* \le \frac{a}{b+k}$$

*for $k$ sufficiently large. Moreover, if $\phi$ satisfies the strong convexity condition (4.1), then there exists $\theta \in (0,1)$ and $c$ such that*

$$\phi(\mathbf{x}_k) - \phi^* \le c\theta^k(\phi(\mathbf{x}_1) - \phi^*)$$

*for every $k$.*

*Proof.* Let $k_i$, $i = 1, 2, \ldots$, denote an increasing sequence of integers with the property that $\phi_j^R \le \phi_j^{\max}$ for $j = k_i$ and $\phi_j^R \le \phi_{j-1}^R$ when $k_i < j < k_{i+1}$. Such a sequence exists since $\phi_k^R \le \max\{\phi_{k-1}^R, \phi_k^{\max}\}$ for each $k$ and (5.2) holds. Moreover, $k_{i+1} - k_i \le L$. Hence, we have

$$\phi_j^R \le \phi_{k_i}^R \le \phi_{k_i}^{\max}, \quad \text{when } k_i \le j < k_{i+1}. \tag{5.3}$$

Let us define

$$\phi_j^{\max+} = \max\{\phi(\mathbf{x}_{j-i} : 0 \le i < \min(j, M+L)\}.$$

Given $j$, choose $k_i$ such that $j \in [k_i, k_{i+1})$. Since $j - k_i < L$, the set of function values maximized to obtain $\phi_{k_i}^{\max}$ is contained in the set of function values maximized to obtain $\phi_j^{\max+}$ and we have

$$\phi_{k_i}^{\max} \le \phi_j^{\max+}. \tag{5.4}$$

Combining (5.3) and (5.4) yields $\phi_j^R \le \phi_j^{\max +}$ for each $j$. In Step 2 of SpaRSA, the iterates are chosen to satisfy the condition

$$\phi(\mathbf{x}_{k+1}) \le \phi_k^R - \sigma\alpha\|\mathbf{x}_{k+1} - \mathbf{x}_k\|^2.$$

It follows that

$$\phi(\mathbf{x}_{k+1}) \le \phi_k^{\max +} - \sigma\alpha\|\mathbf{x}_{k+1} - \mathbf{x}_k\|^2.$$

Hence, the iterates also satisfy the GLL condition, but with memory of length $M + L$ instead of $M$. By Theorem 3.2, the iterates converge at least sublinearly. Moreover, if the strong convexity condition (4.1) holds, then the convergence is R-linear by Theorem 4.1. □

**6. Computational experiments.** In this section, we compare the performance of SpaRSA with the GLL reference function value $\phi_k^{\max}$ and the BB choice for $\alpha_0$ in SpaRSA, to that of an adaptive implementation based on the reference function value $\phi_k^R$ given in the appendix of [15] and a cyclic BB choice for $\alpha_0$. We call this implementation Adaptive SpaRSA. This adaptive choice for $\phi_k^R$ satisfies (R1)–(R3) which ensures convergence in accordance with Theorem 5.1. By a cyclic choice for the BB parameter (see [8, 9, 13, 19]), we mean that $\alpha_0 = \alpha_k^{BB}$ is reused for several iterations. More precisely, for some integer $m \ge 1$ (the cycle length), and for all $k \in ((i-1)m, im]$, the value of $\alpha_0$ at iteration $k$ is given by

$$(\alpha_0)_k = \alpha_{(i-1)m+1}^{BB}.$$

The test problems are associated with applications in the areas of signal processing and image reconstruction. All experiments were carried out on a PC using Matlab 7.6 with a AMD Athlon 64 X2 dual core 3 Ghz processor and 3GB of memory running Windows Vista. Version 2.0 of SpaSRA was obtained from Mário Figueiredo's webpage (http://www.lx.it.pt/∼mtf/SpaRSA/). The code was run with default parameters. Adaptive SpaRSA was written in Matlab with the following parameter values

$$\alpha_{\min} = 10^{-30}, \quad \alpha_{\max} = 10^{30}, \quad \eta = 5, \quad \sigma = 10^{-4}, \quad M = 10.$$

The test problems, such as the basis pursuit denoising problem (1.2), involve a parameter $\tau$. The choice of the cycle length was based on the value of $\tau$:

$$m = 1 \text{ if } \tau \ge 10^{-2}, \text{ otherwise } m = 3.$$

As $\tau$ approaches zero, the optimization problem becomes more ill conditioned and the convergence speed improves when the cycle length is increased.

The stopping condition for both SpaRSA and Adaptive SpaRSA was

$$\alpha_k\|\mathbf{x}_{k+1} - \mathbf{x}_k\|_\infty \le \epsilon,$$

where $\alpha_k$ denotes the final value for $\alpha$ in Step 2 of SpaRSA, $\|\cdot\|_\infty$ is the max-norm, and $\epsilon$ is the error tolerance. This termination condition is suggested by Vandenberghe in [22]. As pointed out earlier, $\mathbf{x}_k$ is a stationary point when $\mathbf{x}_{k+1} = \mathbf{x}_k$. For other stopping criteria, see [16] or [24]. In the following tables, "Ax" denotes the number of times that a vector is multiplied by $\mathbf{A}$ or $\mathbf{A}^\mathsf{T}$, "cpu" is the CPU time in seconds, and "Obj" is the objective function value.

**6.1. $\ell_2 - \ell_1$ problems.** We compare the performance of Adaptive SpaRSA with SpaRSA by solving $\ell_2 - \ell_1$ problems of form (1.2) using the randomly generated data introduced in [17, 24]. The matrix $\mathbf{A}$ is a random $k \times n$ matrix, with $k = 2^8$ and $n = 2^{10}$. The elements of $\mathbf{A}$ are chosen from a Gaussian distribution with mean zero and variance $1/(2n)$. The observed vector is $\mathbf{b} = \mathbf{A}\mathbf{x}_{true} + \mathbf{n}$, where the noise $\mathbf{n}$ is sampled from a Gaussian distribution with mean zero and variance $10^{-4}$. $\mathbf{x}_{true}$ is a vector with 160 randomly placed $\pm 1$ spikes with zeros in the remaining elements. This is a typical sparse signal recovery problem which often arises in compressed sensing [11]. We solved the problem (1.2) corresponding to the error tolerance $10^{-5}$ with different regularization parameters $\tau$ between $10^{-1}$ and $10^{-5}$. Table 6.1 reports the average cpu times (seconds) and the number of matrix-vector multiplications over 10 runs for both the original SpaRSA algorithm and an implementation based on a continuation method (see [16]). The implementations using the continuation method are indicated by "/c" in Table 6.1. These results show that the Adaptive SpaRSA is significantly faster than SpaSRA when not using the continuation technique. The performance gap decreases when the continuation technique is applied. Nonetheless, Adaptive SpaRSA yields better performance.

Figure 6.1 plots error versus the number of matrix-vector multiplication for $\tau = 10^{-4}$ and the implementation without continuation. When the error is large, both algorithm have the same performance. As the error tolerance decreases, the performance of the adaptive algorithm is significantly better than the original implementation.

TABLE 6.1
*Average over 10 runs for $\ell_2 - \ell_1$ problems*

| $\tau$ | 1e-1 | | 1e-2 | | 1e-3 | | 1e-4 | | 1e-5 | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Ax | cpu | Ax | cpu | Ax | cpu | Ax | cpu | Ax | cpu |
| SpaRSA | 65.3 | .07 | 706.4 | .56 | 3467.5 | 2.73 | 8802.9 | 6.86 | 5925.5 | 4.65 |
| Adaptive | 65.4 | .07 | 582.8 | .44 | 1998.8 | 1.58 | 4394.0 | 3.50 | 2911.9 | 2.36 |
| SpaRSA/c | 65.3 | .07 | 626.7 | .48 | 2172.1 | 1.67 | 684.9 | .52 | 474.8 | .36 |
| Adaptive/c | 65.4 | .07 | 569.0 | .44 | 1928.3 | 1.51 | 636.0 | .50 | 453.7 | .34 |

**6.2. Image deblurring problems.** In this subsection, we present results for two image restoration problems based on images referred to as *Resolution* and *Cameraman*. The images are $256 \times 256$ gray scale images; that is, $n = 256^2 = 65536$. The images are blurred by convolution with an $8 \times 8$ blurring mask and normally distributed noise with standard deviation 0.0055 is added to the final signal (see problem 701 in [21]). The image restoration problem has the form (1.2) where $\tau = 0.00005$ and $\mathbf{A} = \mathbf{HW}$ is the composition of the blur matrix and the Haar discrete wavelet transform (DWT) operator. For these test problems, the continuation approach is no faster, and in some cases significantly slower, than the implementation without continuation. Therefore, we solved these test problems without the continuation technique. The results in Table 6.2 again indicate that the adaptive scheme yields much better performance as the error tolerance decreases.

**6.3. Group-separable regularizer.** In this subsection, we examine performance using the group separable regularizers [24] for which

$$\psi(\mathbf{x}) = \tau \sum_{i=1}^{n} \|\mathbf{x}_{[i]}\|_2,$$

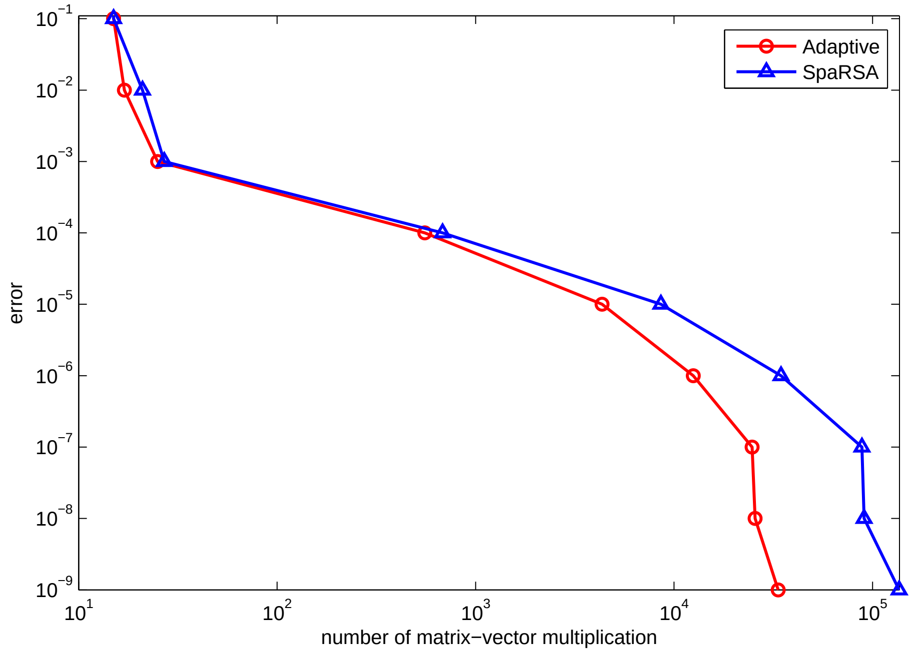


FIG. 6.1. *Number of matrix-vector multiplications versus error*

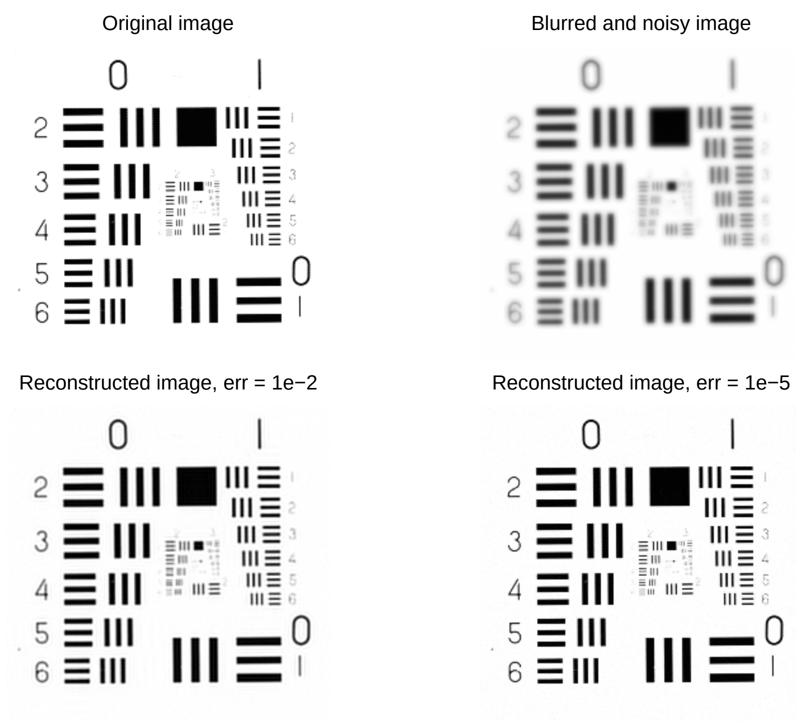


FIG. 6.2. *Deblurring the resolution image*

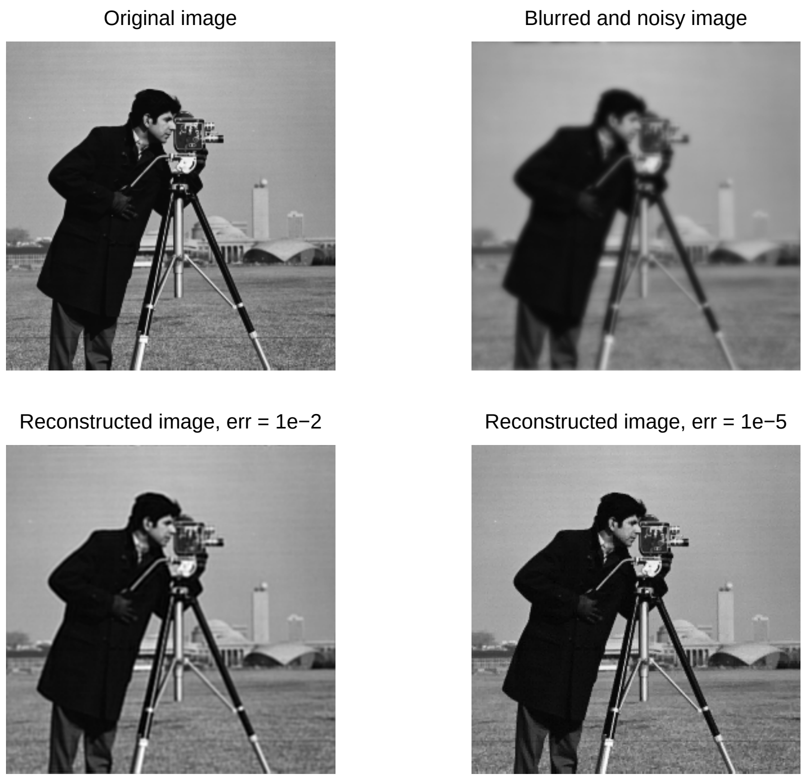


FIG. 6.3. *Deblurring the cameraman image*

where $\mathbf{x}_{[1]}, \mathbf{x}_{[2]}, \ldots, \mathbf{x}_{[m]}$ are $m$ disjoint subvectors of $\mathbf{x}$. The smooth part of $\phi$ can be expressed as $f(\mathbf{x}) = \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2$, where $\mathbf{A} \in \mathbb{R}^{1024\times 4096}$ was obtained by orthonormalizing the rows of a matrix constructed in Subsection 6.1. The true vector $\mathbf{x}_{true}$ has 4096 components divided into $m = 64$ groups of length $l_i = 64$. $\mathbf{x}_{true}$ is generated by randomly choosing 8 groups and filling them with numbers chosen from a Gaussian distribution with zero mean and unit variance, while all other groups are filled with zeros. The target vector is $\mathbf{b} = \mathbf{A}\mathbf{x}_{true} + \mathbf{n}$, where $\mathbf{n}$ is Gaussian noise with mean

TABLE 6.2
*Deblurring images*

| error | | | 1e-2 | | | 1e-3 | | | 1e-4 | | | 1e-5 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Ax | cpu | Obj | Ax | cpu | Obj | Ax | cpu | Obj | Ax | cpu | Obj |
| Resolution | | | | | | | | | | | | |
| SpaRSA | 49 | 2.57 | .4843 | 88 | 4.80 | .3525 | 458 | 24.74 | .2992 | 1679 | 88.27 | .2970 |
| Adaptive | 37 | 1.93 | .5619 | 73 | 4.02 | .3790 | 316 | 17.28 | .2981 | 681 | 35.90 | .2970 |
| Cameraman | | | | | | | | | | | | |
| SpaRSA | 34 | 1.66 | .3491 | 77 | 3.99 | .2181 | 332 | 17.08 | .1880 | 1356 | 69.45 | .1868 |
| Adaptive | 35 | 1.71 | .3380 | 63 | 3.31 | .2232 | 215 | 11.20 | .1880 | 599 | 31.4 | .1868 |

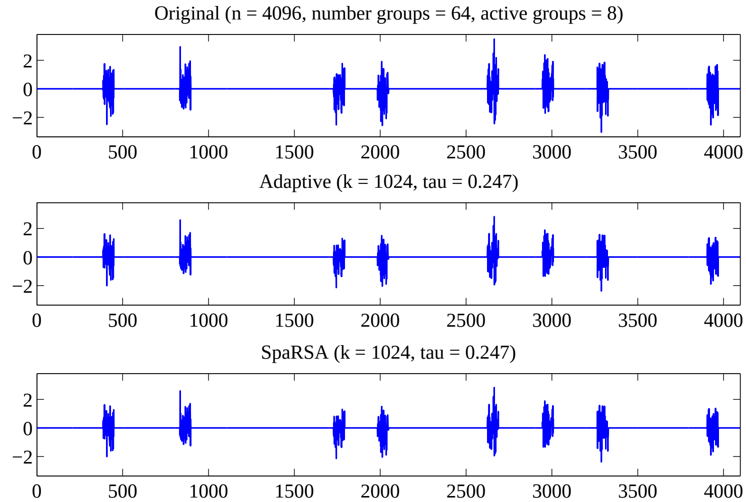


FIG. 6.4. *Group-separable reconstruction*

zero and variance $10^{-4}$. The regularization parameter is chosen as suggested in [24]: $\tau = 0.3\|\mathbf{A}^{\mathsf{T}}\mathbf{b}\|_\infty$. We ran 10 test problems with error tolerance $= 10^{-5}$ and compute the average results. Adaptive SpaRSA solved the test problem in 0.8420 seconds with 67.4 matrix/vector multiplications, while the SpaRSA obtained similar performance: 0.8783 seconds and 69.1 matrix/vector multiplications. Figure 6.4 shows the result obtained by both methods for one sample.

**6.4. Total-variation phantom reconstruction.** In this experiment, the image is the Shepp-Logan phantom of size $256 \times 256$ (see [3, 5]). The objective function was

$$\phi(\mathbf{x}) = \frac{1}{2}\|\mathbf{A}(\mathbf{x}) - \mathbf{b}\|^2 + .01\mathrm{TV}(\mathbf{x})$$

where $\mathbf{A}$ is a $6136 \times 256^2$ matrix corresponding to 6136 locations in the 2D Fourier plane (`masked_FFT` in Matlab). The total variation (TV) regularization is defined as follows

$$\mathrm{TV}(\mathbf{x}) = \sum_i \sqrt{\left(\triangle_i^h \mathbf{x}\right)^2 + \left(\triangle_i^v \mathbf{x}\right)^2}$$

where $\triangle_i^h$ and $\triangle_i^v$ are linear operators corresponding to horizontal and vertical first order differences (see [4]). As seen in Table 6.3, Adaptive SpaRSA was faster than the original SpaRSA when the error tolerance was sufficiently small.

TABLE 6.3
*Total-variation phantom reconstruction*

| error | | | 1e-2 | | | 1e-3 | | | 1e-4 |
|---|---|---|---|---|---|---|---|---|---|
| | Ax | cpu | Obj | Ax | cpu | Obj | Ax | cpu | Obj |
| SpaRSA | 14 | 2.55 | 36.7311 | 143 | 30.06 | 14.7457 | 2877 | 938.25 | 14.1433 |
| Adaptive | 14 | 2.57 | 36.7311 | 136 | 27.32 | 14.6840 | 731 | 185.62 | 14.1730 |

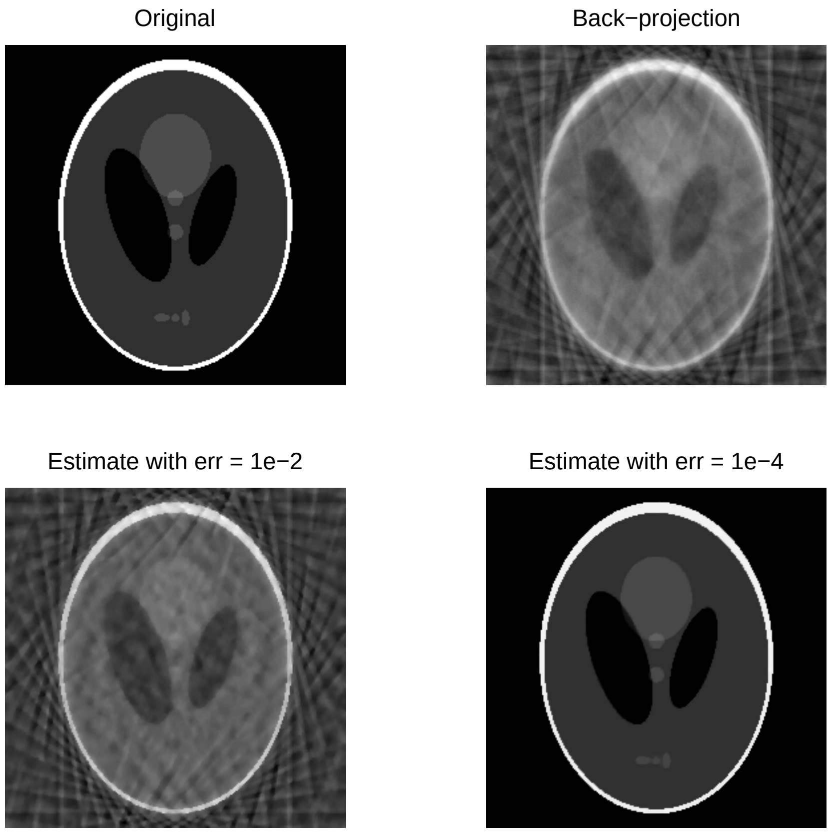


Fig. 6.5. *Phantom reconstruction*

**7. Conclusions.** The convergence properties of the SpaRSA algorithm (Sparse Reconstruction by Separable Approximation) of Wright, Nowak, and Figueiredo [24] are analyzed. We establish sublinear convergence when $\phi$ is convex and the GLL reference function value [14] is employed. When $\phi$ is strongly convex, the convergence is R-linear. For a reference function value which satisfies (R1)–(R3), we prove the existence of a convergent subsequence of iterates that approaches a stationary point. For a slightly stronger version of (R3), given in (5.2), we show that sublinear or linear convergence again hold when $\phi$ is convex or strongly convex respectively. In a series of numerical experiments, it is shown that an Adaptive SpaRSA, based on a relaxed choice of the reference function value and a cyclic BB iteration [9, 15], often yields much faster convergence, especially when the error tolerance is small.

## REFERENCES


[1] J. Barzilai and J. M. Borwein, *Two point step size gradient methods*, IMA J. Numer. Anal., 8 (1988), pp. 141–148.

[2] A. Beck and M. Teboulle, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM Journal on Imaging Sciences, 2 (2009), pp. 183–202.

[3] J. Bioucas-Dias and M. Figueiredo, *Twist: Two-step iterative shrinkage/thresholding algorithm for linear inverse problems.* http://www.lx.it.pt/∼bioucas/TwIST/TwIST.htm.

[4] J. Bioucas-Dias, M. Figueiredo, and J. P. Oliveira, *Total variation-based image deconvolution: a majorization-minimization approach.*, in Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, vol. 2, 2006, pp. 861–864.

[5] E. J. Candès and J. Romberg, *Practical signal recovery from random projections.*, Wavelet Applications in Signal and Image Processing XI, Proc. SPIE Conf., 5914 (2005).

[6] A. Chambolle, R. A. DeVore, N. Y. Lee, and B. J. Lucier, *Nonlinear wavelet image processing: Variational problems, compression, and noise removal through wavelet shrinkage*, IEEE Trans. Image Process., 7 (1998), p. 319335.

[7] S. Chen, D. Donoho, and M. Saunders, *Atomic decomposition by basis pursuit*, SIAM J. Sci. Comput., 20 (1998), pp. 33–61.

[8] Y. H. Dai, *Alternate stepsize gradient method*, Optimization, 52 (2003), pp. 395–415.

[9] Y. H. Dai, W. W. Hager, K. Schittkowski, and H. Zhang, *The cyclic Barzilai-Borwein method for unconstrained optimization*, IMA J. Numer. Anal., 26 (2006), pp. 604–627.

[10] I. Daubechies, M. Defrise, and C. D. Mol, *An iterative thresholding algorithm for linear problems with a sparsity constraint*, Comm. Pure Appl. Math., 57 (2004), pp. 1413–1457.

[11] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright, *Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems*, IEEE Journal on Selected Topics in Signal Processing, 1 (2007), pp. 586–597.

[12] T. Figueiredo and R. D. Nowak, *An EM algorithm for wavelet-based image restoration*, IEEE Trans. Image Process., 12 (2003), p. 906916.

[13] A. Friedlander, J. M. Martínez, B. Molina, and M. Raydan, *Gradient method with retards and generalizations*, SIAM J. Numer. Anal., 36 (1999), pp. 275–289.

[14] L. Grippo, F. Lampariello, and S. Lucidi, *A nonmonotone line search technique for Newton's method*, SIAM J. Numer. Anal., 23 (1986), pp. 707–716.

[15] W. W. Hager and H. Zhang, *A new active set algorithm for box constrained optimization*, SIAM J. Optim., 17 (2006), pp. 526–557.

[16] E. Hale, W. Yin, and Y. Zhang, *A fixed-point continuation method for $\ell_1$-regularized minimization with applications to compressed sensing*, tech. report, Rice University, July 2007.

[17] S.-J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky, *An interior-point method for large-scale $\ell_1$-regularized least squares*, IEEE Journal on Selected Topics in Signal Processing, 1 (2007), pp. 606–617.

[18] Y. Nesterov, *Gradient methods for minimizing composite objective function*, CORE Discussion Papers 2007/76, Universit catholique de Louvain, Center for Operations Research and Econometrics (CORE), Sept. 2007.

[19] M. Raydan and B. F. Svaiter, *Relaxed steepest descent and Cauchy-Barzilai-Borwein method*, Comput. Optim. Appl., 21 (2002), pp. 155–167.

[20] R. T. Rockafellar, *Convex analysis*, Princeton Univ. Press, 1970.

[21] E. van den Berg, M. P. Friedlander, G. Hennenfent, F. J. Herrmann, R. Saab, and O. Yilmaz, *Algorithm 890: Sparco: A testing framework for sparse reconstruction*, ACM Trans. Math. Softw., 35 (2009), pp. 1–16.

[22] L. Vandenberghe, *Gradient methods for nonsmooth problems (lecture note - spring 2009).* http://www.ee.ucla.edu/∼vandenbe/ee236c.html.

[23] C. Vonesch and M. Unser, *Fast iterative thresholding algorithm for wavelet-regularized deconvolution*, in Proceedings of the SPIE Optics and Photonics 2007 Conference on Mathematical Methods: Wavelet XII, vol. 6701, San Diego, CA, 2007, pp. 1–5.

[24] S. J. Wright, R. D. Nowak, and M. A. T. Figueiredo, *Sparse reconstruction by separable approximation*, IEEE Trans. Signal Process., 57 (2009), pp. 2479–2493.